JOURNAL OF BACTERIOLOGY, Aug. 1997, p. 4768–4777 Vol. 179, No. 15
0021-9193/97/$04.00+0

Putative Evolutionary Origin of Plasmids Carrying the Genes Involved in Leucine Biosynthesis in *Buchnera aphidicola* (Endosymbiont of Aphids)

ROELAND C. H. J. VAN HAM,† ANDRÉS MOYA, AND AMPARO LATORRE*

Department of Genetics, University of Valencia, 46100 Burjassot, Valencia, Spain

Received 17 March 1997/Accepted 27 May 1997

An 8.5-kb plasmid encoding genes (*leuABCD*) involved in leucine biosynthesis and a small plasmid of 1.74 kb of yet unknown function were found in the intracellular symbiont, *Buchnera aphidicola*, of two divergent aphid species, *Thelaxes suberi* and *Tetraneura caerulescens*, respectively. The *leuABCD*-carrying plasmid (pBTs1) was amplified from total aphid DNA by inverse long PCR, using outwardly oriented oligonucleotide primers specific to *leuA*. The resulting 8.2-kb PCR fragment as well as the 1.74-kb plasmid (pBTc1) were cloned and sequenced. pBTs1 differed from a previously described *B. aphidicola* plasmid (pRPE) of the aphid *Rhopalosiphum padi* by the presence of a small heat shock gene (*ibp*) and in the order of the *leuABCD* and *repA* genes. Comparison of both leucine plasmids to the small plasmid pBTc1 revealed extensive similarity with respect to putative replication functions as well as in the presence of a highly conserved open reading frame that was found to be homologous to *Escherichia coli* YqhA and *Haemophilus influenzae* HI0507 and which may encode an integral membrane protein. The three *B. aphidicola* plasmids most likely evolved from a common ancestral replicon, which in turn may be distantly related to IncFII plasmids. Phylogenetic affiliations of the *B. aphidicola* strains of the two aphid species were assessed by sequencing of their 16S rRNA genes. Evaluation of the distribution of the *leuABCD*-encoding plasmids within a phylogenetic framework suggests independent origins for pBTs1 and pRPE from an ancestral replicon resembling pBTc1. The implications for symbiotic essential amino acid biosynthesis and provisioning are discussed.

Nearly all aphids maintain an endosymbiotic association with *Buchnera aphidicola*, a member of the class *Proteobacteria* that is closely related to *Escherichia coli* (47, 62). The bacteria are harbored in specialized cells of organ-like structures (bacteriomes) in the hemocoel of the aphid. Their transmission is tightly regulated and occurs via the maternal lineage through infection of eggs or embryos (9, 29). *B. aphidicola* has never been cultured and has never been found outside aphids, while the latter, when stripped of the bacteria by treatment with antibiotics, grow slowly and are unable to reproduce (17, 19, 58). The association is therefore considered obligate for both organisms. Phylogenetic analyses of *B. aphidicola* 16S rDNA sequences and aphid morphology have indicated that the association had a single origin, dated about 200 to 250 million years ago, and that host and endosymbiont lineages have diverged strictly in parallel (44, 49). Genetic studies suggest that *B. aphidicola*, in spite of its well-established and sheltered relationship within aphids, has retained many characteristics of free-living bacteria (6).

B. aphidicola is thought to substantially contribute to the ability of aphids to thrive on plant sap, a diet of unbalanced composition that is notably poor in nitrogenous compounds (15, 18, 31, 54). Various specific nutritional roles have been proposed, but there is strong experimental evidence only for bacterial supply of essential amino acids to the aphid (16, 20, 21, 58). Tryptophan biosynthesis and provisioning have been studied in most detail and have rendered a coherent body of physiological, metabolic, and molecular genetic data. Douglas and Prosser (20) showed that symbiotic pea aphids (*Acyrthosiphon pisum*, Aphididae) could be maintained for several generations on artificial diets, either supplemented with or lacking tryptophan. In contrast, aphids stripped of their bacteria (aposymbiotic aphids) by chlortetracycline treatment rarely reached adulthood when tryptophan was lacking from their diet but showed statistically significant survival when fed on a complete diet. In addition, tryptophan synthetase activity was detected in symbiotic aphids and isolated bacteria but not in aposymbiotic aphids (20). In line with these findings, Lai et al. (36) discovered the transfer of *trpEG*, encoding the key regulatory enzyme in tryptophan biosynthesis, anthranilate synthase, to a plasmid in *B. aphidicola* of the aphid *Schizaphis graminum* (Aphididae). It was shown that this rearrangement resulted in a 14- to 15-fold amplification of *trpEG* relative to the other genes of the pathway, which still reside in the *B. aphidicola* chromosome (46). The observations have been interpreted in terms of adaptation of *B. aphidicola* to its symbiotic environment, endowing it with the capacity to overproduce tryptophan, which is supplied to the aphid host.

Recently, Bracho et al. (8) discovered a plasmid in *B. aphidicola* from the aphid *Rhopalosiphum padi* (Aphididae) that carries all the genes (*leuABCD*) of the biosynthetic pathway leading to the immediate precursor (α-ketoisocaproate) of the essential amino acid leucine. This genetic organization, similar to that of *trpEG*, has been interpreted as a capacity of the bacteria to overproduce leucine. To further assess the significance and dynamics of genetic modifications in the leucine biosynthetic pathway, we are characterizing the structural organization of the gene cluster in *B. aphidicola* from several major lineages of aphids. Here, we present the finding and comparative sequence analysis of (i) an 8.5-kb plasmid carrying the leucine gene cluster in *B. aphidicola* from the aphid

* Corresponding author. Mailing address: Department of Genetics, University of Valencia, c/Dr. Moliner 50, 46100 Burjassot, Valencia, Spain. Phone: 34-6-3864505. Fax: 34-6-3983029. E-mail: ampazo.latorre@euv.es.

† Present address: Department of Biology, University of York, York, United Kingdom.

Thelaxes suberi (Thelaxidae) and (ii) a 1.74-kb plasmid in *B. aphidicola* from the aphid *Tetraneura caerulescens* (Pemphigidae) that bear resemblance to the leucine plasmids in putative replication functions as well as in the presence of a highly conserved open reading frame. Plasmids in *B. aphidicola* from aphids other than the Aphididae have not previously been described. Using a phylogenetic framework based on 16S rDNA sequence comparisons, we hypothesize that an ancestral replicon, probably resembling the smaller plasmid, has been specifically recruited for amplification of the leucine gene cluster more than once during the coevolution of *Buchnera* and aphids.

MATERIALS AND METHODS

Aphid material and *B. aphidicola*. Plant material carrying aphids was collected in the field and kept at 4°C until further manipulations of the animals. *Thelaxes suberi* was collected from a small population on oak trees in Alcublas, Spain, and *Tetraneura caerulescens* was sampled several times from a large gall-forming population on elm trees in Bugarra, Spain. In accordance with the practice introduced by Baumann et al. (6), *B. aphidicola* isolates from different aphid species are designated by *B* followed by the name of the aphid host, e.g., *B*(*Thelaxes suberi*) refers to *B. aphidicola* from the aphid *Thelaxes suberi*.

DNA preparation. Isolation of total aphid DNA, which contains both host DNA and endosymbiont DNA, was the preferred method if only limited numbers of individuals (≤0.5 g [fresh weight]) could be collected. Total DNAs from *Thelaxes suberi* and *Tetraneura caerulescens* were isolated as described previously (38). *B. aphidicola* was isolated from *Tetraneura caerulescens* according to the protocol of Harrison et al. (27). Genomic DNA of *B. aphidicola* was isolated by using the CTAB/NaCl extraction protocol (3), and plasmid DNA from *B*(*Tetraneura caerulescens*) was prepared by using Qiagen-tip 20 gravity flow columns (Qiagen).

General methods. DNA techniques were performed according to standard methods (57). PCR products and restriction endonuclease digests of DNA preparations were separated by gel electrophoresis on 0.8 or 1.4% agarose gels and transferred to nylon membranes (Hybond N+; Amersham). Labelling of probes and Southern hybridization were performed by using the enhanced chemiluminescence (ECL) system (Amersham). Probe il7.5 was prepared from a purified long PCR product and covers nearly the entire sequence of pRPE, the *leuABCD*-carrying plasmid from *B*(*R. padi*) (8). Probes specific to single genes were prepared from purified PCR products, which were obtained with sets of primers specific to pRPE and *R. padi* DNA as templates. Hybridizations were carried out at 42°C overnight in ECL Gold hybridization buffer supplemented with 5% blocking agent and 0.5 M NaCl in a rotating hybridization oven. After hybridization, membranes were washed twice for 20 min at 42°C in 6 M urea–0.4% sodium dodecyl sulfate–0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), followed by two washes for 5 min at room temperature in 2× SSC. Chemiluminescence was detected on autoradiography films (Kodak) after signal generation with ECL detection reagents.

Inverse long PCR. In screening the structural organization of the leucine gene cluster in *B. aphidicola* strains from major lineages of aphids, we used a combination of structural PCR assays, Southern hybridization, and nucleotide sequencing of selected products. The structural PCR assays included application of the inverse long PCR technique (51). Degenerate primers complementary to the *leuA* gene and outwardly oriented were designed, based on comparison of nucleotide sequences from *B*(*R. padi*), *E. coli*, and *Salmonella typhimurium* and taking into account the codon usage of *B. aphidicola* (12). 5′ extensions with restriction sites for *Bam*HI and *Pst*I or *Sal*I were made, resulting in the oligonucleotides leuA.dl3 (5′ CGG ATC CGT CGA CAR ACT WGC TTG WAR WGC TTG TTC WCC ATC 3′) and leuA.du2 (5′ CGG ATC CTG CAG GAT GAT GTW GAA TTT TCW TGY GAR GAY GC 3′), corresponding to positions 3181 to 3153 and 3513 to 3541 of the pRPE sequence from *B*(*R. padi*), respectively (GenBank accession no. X71612). With total aphid or total *B. aphidicola* DNA as a template, these primers are expected to amplify by long PCR either (i) a limited stretch of (chromosomal or plasmid) DNA that carries a duplicated *leuA* gene or (ii) an entire plasmid that carries a single copy of the *leuA* gene. Amplification of natural plasmid DNA from *B*(*Thelaxes suberi*) carrying the leucine gene cluster was done with 100 ng of total aphid DNA as a template. Cycling was performed on a GeneAmp 2400 System (Perkin-Elmer) using the Expand Long PCR kit (Boehringer Mannheim). Automated "hotstart" PCR was done with Ampliwax (Perkin-Elmer) according to the manufacturer's instructions. Cycling conditions were as follows: 92°C for 2 min; 10 cycles consisting of 92°C for 10 s, 62°C for 30 s, and 68°C for 10 min; 20 cycles identical to the previous round except for a 20-s autoextension of the extension time per cycle; and a final 7 min at 68°C.

16S rDNA amplification. The 16S rRNA genes from *B*(*Thelaxes suberi*) and *B*(*Tetraneura caerulescens*) were amplified by PCR using *Taq* DNA polymerase (Pharmacia) and oligonucleotide primers corresponding to *E. coli* rDNA nucleotides 8 to 30 (5′ AGA GTT TGA TCA TGG CTC AGA TTG 3′) and 1507 to 1484 (5′ TAC CTT GTT ACG ACT TCA CCC CAG 3′). Cycling was performed with the GeneAmp 2400 System and consisted of 95°C for 5 min; 30 cycles of 95°C for 10 s, 65°C for 30 s, and 72°C for 2 min; and a final 7 min at 72°C.

Cloning and sequencing. A long PCR product of 8.2 kb, obtained from amplification with primers leuA.dl3 and leuA.du2 and total DNA of *Thelaxes suberi* as a template, was purified by using the GeneClean II kit (Bio 101, La Jolla, Calif.) and digested with *Bam*HI. Two resulting fragments of 4.2 and 4.0 kb were cloned into the vector pUC18. Natural plasmid DNA isolated from *B*(*Tetraneura caerulescens*) was digested with *Eco*RI and shotgun cloned into the vector pBluescript II SK (Stratagene). One of seven recombinant clones carrying an insert of 1.74 kb was selected for further analysis after demonstration by restriction endonuclease digestion that all clones were identical. 16S rDNA PCR products were purified with GeneClean and cloned into a T vector prepared according to the method of Marchuk et al. (42) from *Eco*RV-digested pBluescript II SK. Nested deletions of cloned fragments were generated in both directions by using the Nested Deletions kit (Pharmacia). Plasmid DNAs from clones with suitable insert sizes were purified by using Qiagen-tip 20 gravity flow columns (Qiagen). Nucleotide sequencing was performed on both strands with the AmpliTaqF Dye Deoxy Terminator Cycle Sequencing kit (Perkin-Elmer) and the 373 automated DNA sequencer as recommended by the manufacturer.

Computer analysis of the DNA sequences. DNA sequence data were assembled and analyzed with the Genetics Computer Group (GCG) program package 8.0 for the VAX/VMS (14). BLASTP searches (2) were done at the network server of the NCBI. Hydrophobicity analysis and transmembrane predictions (30, 34) were done at the ExPASy molecular biology server of the Geneva University Hospital and the University of Geneva and the server of the Bioinformatics Group at the Swiss Institute for Experimental Cancer Research (ISREC). Phylogenetic analysis of amino acid sequences was performed with PHYLIP version 3.5 (23). 16S rDNA sequences were aligned at the network server of the Ribosomal Database Project (RDP) (41), among-site rate variation was assessed with the PAML program version 1.2 (66), and categorized rate estimates were implemented in phylogenetic analysis of the sequences using fastDNAml version 1.0 for the VAX/VMS (53).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper have been deposited in the GenBank/EMBL database under accession numbers Y11966 and Y11972 [pBTs1 and 16S rDNA sequences from *B*(*Thelaxes suberi*), respectively] and Y11974 and Y11973 [pBTc1 and 16S rDNA sequences from *B*(*Tetraneura caerulescens*), respectively].

RESULTS

Identification of pBTs1. Only a limited amount of *Thelaxes suberi* could be collected. Total DNA extraction, containing aphid and *B. aphidicola* genomic and plasmid DNAs, was therefore used to maximize yield. Southern blot analysis of undigested and digested total DNA of *Thelaxes suberi* indicated that one or more of the *leuABCD* genes of a bacterial endosymbiont map to a circular plasmid, because (i) different DNA forms were detected in undigested samples, which most likely correspond to covalently closed circular, open circular, and linear forms of a circular plasmid, and (ii) the sum of fragment sizes for different restriction enzymes was invariably about 8.5 kb (Fig. 1A). The fact that no additional fragments were detected suggests the absence of a chromosomal copy of *leuABCD*. In accordance with the estimated size of 8.5 kb for the putative plasmid, a long PCR product of approximately 8.2 kb was obtained after amplification with outwardly oriented primers located 331 nucleotides apart within *leuA* and with total DNA of *Thelaxes suberi* as a template (Fig. 1B). Restriction fragment patterns of the PCR product were compatible with those obtained in the initial hybridization. By analogy with the organization of the leucine gene cluster previously observed in *B*(*R. padi*) (8) and on the basis of analyses of amino acid sequences presented below, it is most likely that a plasmid carrying genes involved in leucine biosynthesis is contained in *B*(*Thelaxes suberi*). The plasmid was designated pBTs1. The corresponding long PCR product was digested with *Bam*HI, and the two resulting fragments of 4.2 and 4.0 kb were cloned into the vector pUC18.

Structural features and open reading frame analysis of pBTs1. The DNA sequence of the entire long PCR product was determined for both strands. The total length of the fragment is 8,106 bp, excluding two primer sequences of 41 and 42

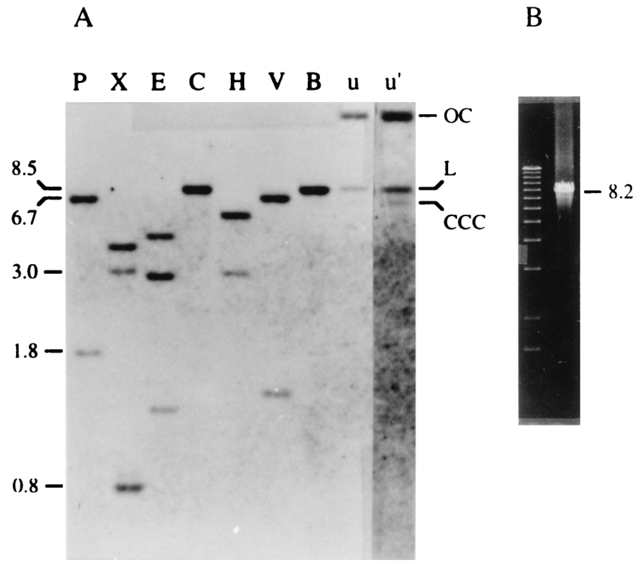

FIG. 1. Analysis of *Thelaxes suberi* DNA. (A) Southern blot of restriction enzyme digests and undigested total DNA hybridized with probe il7.5. P, *Pst*I; X, *Xba*I; E, *Eco*RI; C, *Cla*I; H, *Hin*dIII; V, *Eco*RV; B, *Bam*HI. Lanes u and u′, are identical, but u′ was obtained after longer exposure of the membrane to a second autoradiography film. The different forms of the putative plasmid are open circular (OC), linear (L), and covalently closed circular (CCC). (B) Long PCR product obtained after amplification with outwardly oriented primers located 331 nucleotides apart within *leuA* using total DNA as a template. A 1-kb DNA ladder (Gibco-BRL) was used as a molecular weight marker.

nucleotides. Relative to pRPE, a stretch of 389 bp internal to *leuA* was not determined for pBTs1, due to the location of the primers used for inverse long PCR. The estimated total length of the plasmid is therefore 8.5 kb. A physical map of the sequence is presented in Fig. 2A. For comparison, the map of pRPE from *B*(*R. padi*) is included. The GC content of pBTs1 (25.4%) was similar to that of pRPE (26%) and slightly lower than that of *B*(*Acyrthosiphon pisum*) genomic DNA (28 to 30%) (32). Noncoding regions were particularly AT rich (87.1%) and, most likely due to this composition, were found to contain numerous small inverted and direct repeats. One long direct repeat was found between *ibp* and ORF1, which are identified below. It comprises a perfectly repeated sequence of 52 bp, the two copies of which are spaced by 48 bp. It is unknown whether the repeat is functional or represents a relic from a recombination event.

A search for reading frames of at least 50 codons revealed eight potentially coding sequences. All of these started with ATG and ended with TAA, which is the preferred stop codon in *B. aphidicola* genes (12). Together, the eight open reading frames occupy 85% of the nucleotide sequence of pBTs1. With the exception of *ibp*, all open reading frames are on the same strand. Percentages of similarity and identity between the predicted amino acid sequences and relevant homologs are listed in Table 1.

In agreement with the gene content of pRPE from *B*(*R. padi*), four open reading frames were homologous to *leuA*, *-B*, *-C*, and *-D*, the structural genes that constitute the leucine operon in various bacteria. A remarkable difference between pBTs1 and pRPE pertains to the order of the genes within the cluster. On pRPE, *leuA*, *-B*, *-C*, and *-D* are contiguous in this order, with intergenic spacers of 24, 2, and 2 bp, respectively. This organization is identical to the gene order of the leucine operon in *E. coli* and *S. typhimurium* (61) and *Lactococcus lactis* (26). On pBTs1, in contrast, *leuA* and *leuB* are separated by four open reading frames, yielding an ordered gene cluster of *leuBCDA* with intergenic spacers of 2, 36, and 88 bp, respectively.

BLASTP database homology searches for the putative product of ORF1 yielded high scores (BLASTP probability scores, $\leq 7.9 \times 10^{-40}$) with ORF1 from pRPE, YqhA (ORF_f164) from *E. coli*, and HI0507 from *Haemophilus influenzae*. In addition, a homologous sequence was found on a small plas-

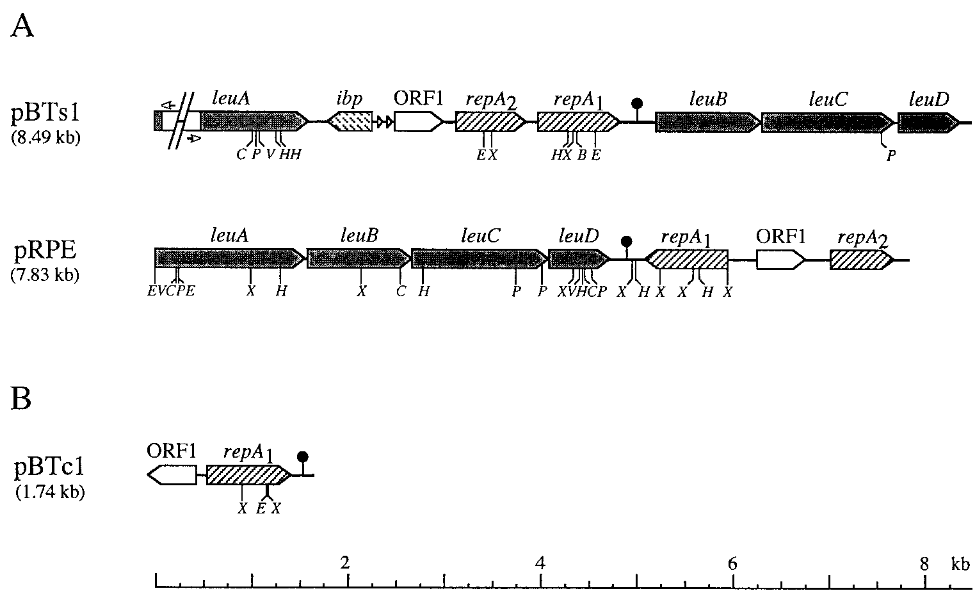

FIG. 2. Linearized physical maps of *B. aphidicola* plasmids. (A) pBTs1 from *B*(*Thelaxes suberi*). The interrupted open box within *leuA* represents the part of the plasmid that has not been sequenced and corresponds to a stretch of 389 nucleotides relative to *leuA* of pRPE from *B*(*R. padi*) (8). The map of pRPE is included for comparison. (B) pBTc1 from *B*(*Tetraneura caerulescens*). Arrows, oligonucleotide primers leuA.dl3 and leuA.du2 used for inverse long PCR; triangles, direct repeats; circles, regions that might contain the origin of replication. Recognition sites of the following enzymes are shown below each map: *Pst*I (*P*), *Xba*I (*X*), *Eco*RI (*E*), *Cla*I (*C*), *Hin*dIII (*H*), *Eco*RV (*V*), and *Bam*HI (*B*).

TABLE 1. Analysis of genes, open reading frames, and potential ribosome-binding sites of pBTs1 and pBTc1[a]

ORF (amino acids), potential RBS, and homologous proteins	% Amino acids	
	Similar	Identical
pBTs1		
leuA (388[b]), TATTTAGAGGGAATT**ATG**		
pRPE LeuA	81.1	67.5
E. coli LeuA	77.5	62.3
leuB (364), TAATGGAGGATAAAA**ATG**		
pRPE LeuB	77.7	60.9
E. coli LeuB	73.5	56.7
leuC (466), AAAAAAGGCTTAAAA**ATG**		
pRPE LeuC	77.3	66.1
E. coli LeuC	78.4	65.5
leuD (205), GTTAATGTTAAAATT**ATG**		
pRPE LeuD	70.1	52.5
E. coli LeuD	66.2	53.7
repA$_1$ (283), TTATTATCATAAATA**ATG**		
pBTs1 RepA$_2$	63.4	49.2
pRPE RepA$_1$	74.6	59.7
pRPE RepA$_2$	68.0	53.6
pBTc1 RepA$_1$	68.0	55.6
R100 RepA	53.7	27.0
repA$_2$ (246), TAGAATAAAAATACT**ATG**		
pRPE RepA$_1$	62.4	48.7
pRPE RepA$_2$	58.6	46.8
pBTc1 RepA$_1$	60.8	50.0
R100 RepA	50.9	23.9
ORF1 (165), TAATAAGGTAATAAA**ATG**		
pRPE ORF1	89.7	81.2
pBTc1 ORF1	85.4	65.2
E. coli YqhA (ORF_f164)	81.8	59.4
H. influenzae HI0507	62.7	34.9
ibp (153), ATGAAGGAGTAATTT**ATG**		
E. coli IbpA	58.4	38.0
E. coli IbpB	58.1	36.8
pBTc1		
ORF1 (178), GAGGGGTTTTTAAGT**ATG**		
pRPE ORF1	87.3	70.3
E. coli YqhA (ORF_f164)	79.8	58.3
H. influenzae HI0507	62.9	35.3
repA$_1$ (289), TTTAGAGGAAAAAAT**ATG**		
pRPE RepA$_1$	67.6	55.6
pRPE RepA$_2$	62.8	47.6
R100 RepA	48.6	21.4

[a] ORF, open reading frame; RBS, ribosome-binding site.

[b] A sequence internal to *leuA*, which is flanked by oligonucleotide primers leuA.dl3 and leuA.du2, which were used for inverse PCR, was not determined. Relative to the length of leuA from pRPE, only the expected number of amino acids can be given.

mid from *B*(*Tetraneura caerulescens*) that is described hereafter. ORF1 is the most conserved open reading frame identified on the present group of plasmids (Table 1). A multiple alignment of the inferred amino acid sequences is shown in Fig. 3A. Hydropathy analysis and prediction of transmembrane segments (30, 34) strongly suggested the presence of at least three transmembrane helices in each of the putative proteins. The hydropathy plot for the product of ORF1 from pBTs1 along with the predicted transmembrane segments is shown in Fig. 4. Plots for the other proteins were highly similar, and their potential transmembrane regions are indicated in Fig. 3A.

The inferred amino acid sequence of an open reading frame on the opposite strand relative to all other open reading frames yielded significant similarity (BLASTP probability scores, $\leq 6.7 \times 10^{-14}$ with various proteins that belong to the hsp20 family of small heat shock proteins (10). The highest similarities were found with IbpA and IbpB from *E. coli* (Table 1), also referred to as HslT and HslS (11) and GspA from *Legionella pneumophila*. In accordance with the nomenclature of Allen et al. (1), the gene is designated *ibp*. The IbpA and IbpB proteins have been shown to constitute major fractions of inclusion bodies formed during high-level expression of heterologous proteins in *E. coli* (1). A composite alignment based on three pairwise alignments of the individual proteins against the profile (kindly provided by J. Leunissen and W. de Jong) of the homologous C-terminal domain of the hsp20 family is presented in Fig. 3B. The presence of *ibp* on pBTs1 represents the only difference in gene content with respect to pRPE.

Two open reading frames were found to be highly similar to each other and to two copies of the same gene located on pRPE. Based on significant sequence similarity, these genes have previously been postulated (8) to code for a homolog of the RepA replication initiation protein of plasmids belonging to the IncFII incompatibility group (65). The two variants of the gene are hereafter referred to as *repA*$_1$ and *repA*$_2$. The *repA*$_1$ variant on each plasmid is expected to encode a protein that is on average 37 amino acids longer than that of *repA*$_2$. In contrast to pRPE, where the two genes occur in an inverted orientation and are separated by ORF1, *repA*$_1$ and *repA*$_2$ of pBTs1 are contiguous and transcribed in the same direction. A multiple alignment of related RepA sequences is presented in Fig. 3C. Both amino acid substitutions and insertions/deletions that account for the difference in length between the two *repA* variants are concentrated between positions 175 and 240 of the multiple alignment.

Identification of pBTc1. During the course of our survey of the organization of the leucine gene cluster among divergent lineages of *B. aphidicola*, we observed a chromosomal location of the gene cluster in *B. aphidicola* from the aphid *Tetraneura caerulescens*. Restriction fragment patterns of *B*(*Tetraneura caerulescens*) genomic DNA obtained after hybridization with probe il7.5, which covers nearly the entire sequence of pRPE, are shown in Fig. 5A. In addition to the fragments that were shown to carry the leucine genes (63) and which yielded strong hybridization signals, each digest revealed either one or two bands of much weaker signal. For the digests in which two bands were detected, these migrated at approximately 1.25 and 1.85 kb. *Eco*RI- and *Xba*I-digested DNAs revealed single bands migrating at 1.74 and 1.48 kb, respectively. Subsequent hybridizations with probes specific to single genes confirmed the suspicion that these weakly hybridizing fragments contained sequences homologous to *repA* and ORF1 (data not shown). These findings suggested the presence of a *repA*-associated replicon in *B*(*Tetraneura caerulescens*) that is not linked to the leucine gene cluster. To confirm this assumption, *B. aphidicola* was isolated from freshly collected aphids and a small-scale preparation of plasmid DNA was made by using an alkaline lysis method (57) and purification on Qiagen gravity flow columns. Digested and undigested plasmid DNAs were electrophoresed in an agarose gel, blotted, and also hybridized with a *repA*-specific probe. Fragments of the same size as in the previous hybridizations were detected (Fig. 5B). In undigested as well as in several restriction enzyme digests of the plasmid

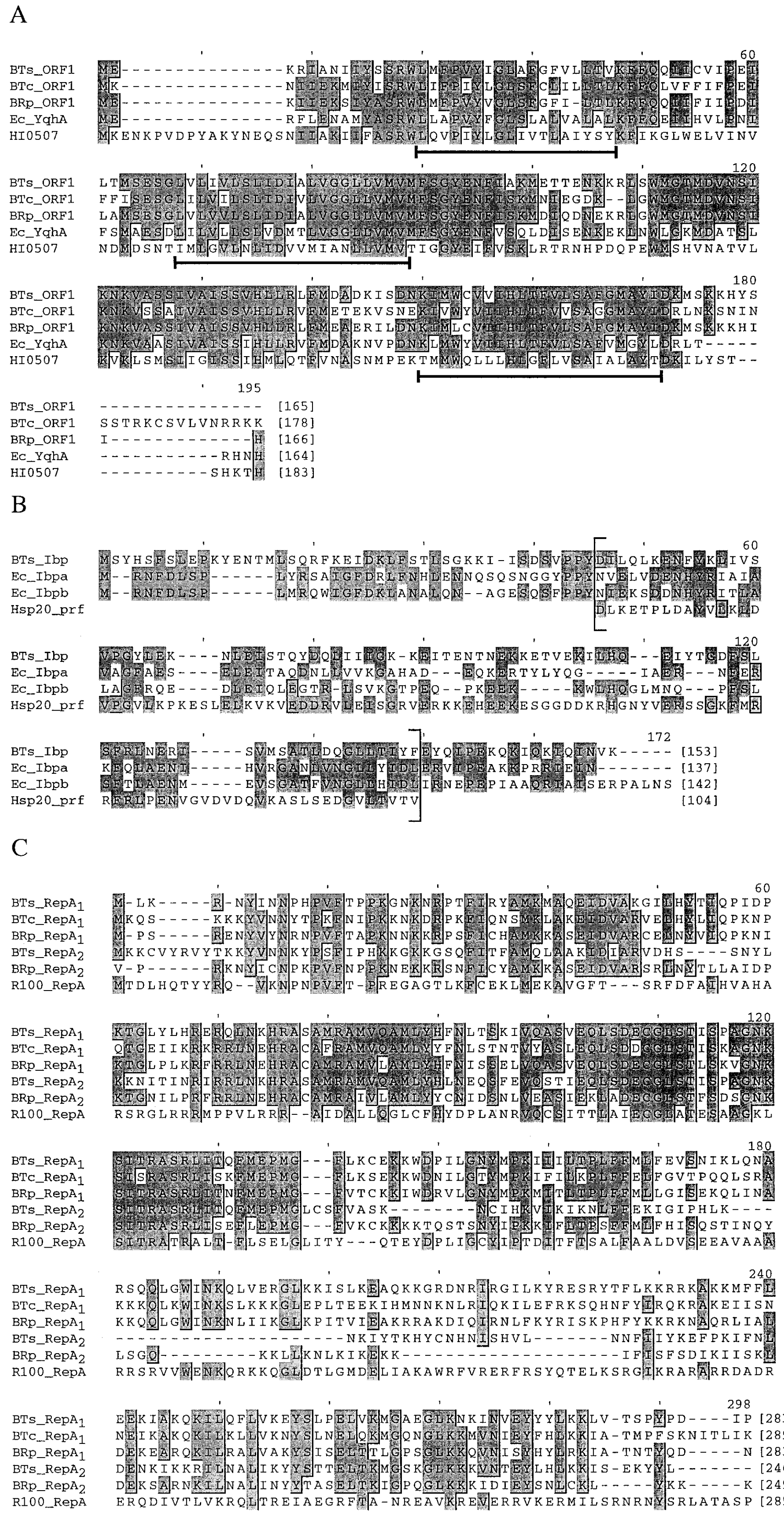

FIG. 3. Multiple alignments of amino acid sequences of putative proteins of *B. aphidicola* plasmids pBTs1 (BTs), pBTc1 (BTc), and pRPE (BRp) and relevant homologous proteins of *E. coli* (Ec) or *H. influenzae* (HI0507) or plasmid R100. Alignment was performed with the Clustal algorithm, unless otherwise stated. (A) ORF1-like sequences. Residues identical in at least three sequences are shaded; bars indicate approximate locations of predicted transmembrane segments. (B) Compilation of pairwise alignments of small heat shock proteins against the profile (prf) of the homologous C-terminal domain of the hsp20 family (bracketed). Residues identical in at least two sequences are shaded. Alignment was performed by using the PROFILEGAP program of the GCG package. (C) RepA-like sequences. Residues identical in at least four sequences are shaded.

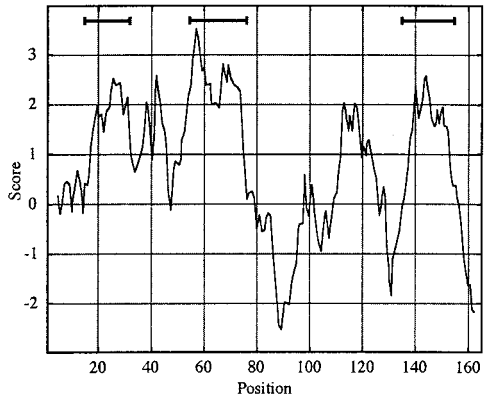

FIG. 4. Hydropathy plot for the putative product of ORF1 from plasmid pBTs1. Bars indicate predicted transmembrane segments. The analyses were performed with the Tmpred program (30), available through the network server at the ISREC.

DNA, a blurred band migrating at around 1.25 kb and a sharp band at 1.85 kb were observed, which most likely correspond to the supercoiled and open circular forms of a circular plasmid of approximately 1.74 kb, respectively. Detection of these bands in restriction enzyme digests was expected to be due to the absence of sites for the respective enzymes on the putative plasmid. For further characterization, the natural plasmid DNA preparation was digested with *Eco*RI, assumed to linearize the molecules, and shotgun cloned into the vector pBluescript II SK. Restriction enzyme analysis of seven recombinants carrying an insert of 1.74 kb showed all the clones to be identical, yielding patterns that were compatible with the initial hybridizations. One clone was completely sequenced on both strands. The single *Eco*RI site was found to be located within a complete *repA* gene, confirming the circular nature of this plasmid, which was designated pBTc1.

Structural features and open reading frame analysis of pBTc1. The GC content of the plasmid was 21.2%, which is the lowest value for a plasmid of *B. aphidicola* described to date. A search for open reading frames revealed the presence of two potentially coding sequences. In agreement with the hybridizations, these were identified as ORF1 and *repA*$_1$ (Fig. 3A and C). A physical map of pBTc1 is presented in Fig. 2B. ORF1 showed high similarity with the homologs from pRPE and pBTs1 (Table 1). In contrast to these plasmids, however, only one copy of the *repA* gene is present on pBTc1.

Potential regulatory sequences on pBTs1 and pBTc1. Potential ribosome-binding sites upstream of initiation codons of genes and open reading frames which are complementary to the 3′ end of *B. aphidicola* 16S rRNA (3′-AUUCCUCCACU U . . . -5′) (48) are listed in Table 1. Inspection of intergenic regions of both plasmids where promoters might be expected did not reveal the presence of sequences closely similar to the consensus *E. coli* σ^{70} promoter sequence, which consists of a −35 (TTGACA) element and a −10 (TATAAT) element separated by 17 ± 2 nucleotides (39). The best match was observed for a region preceding *leuB* on pBTs1, where the sequences TTTACA and TAAAAT were found 17 nucleotides apart, at 138 and 115 bp upstream of the start codon, respectively.

The apparent absence of −10 and −35 regions that closely resemble the eubacterial consensus sequence is a recurrent observation made in sequence analysis studies of *B. aphidicola* DNA (35, 37, 46). Although sequences resembling the −10 region are ubiquitous in *B. aphidicola* DNA, obviously due to its very high A+T content, the corresponding GC-richer elements resembling the −35 sequence at a distance of 17 ± 2 nucleotides are rarely found. Notable exceptions are the regions upstream of the 16S and 23S rRNA-encoding genes, which, in *B. aphidicola* isolates from different aphids, have been found to contain highly conserved −35 sequences that perfectly match the eubacterial consensus sequence (48, 55).

In *E. coli*, expression of the *ibpAB* operon has been shown to be strongly heat inducible (1) and the region upstream of *ibpA* was found to contain promoter elements for Eσ^{32}, the RNA polymerase holoenzyme responsible for transcription of heat shock genes. Nucleotide sequence analysis of the *B. aphidicola* GroESL operon (52) has revealed the presence of −10 and −35 sequences that strongly resembled the *E. coli* consensus sequences recognized by Eσ^{32} (13), indicating that such elements are potentially conserved in *Buchnera*. The region upstream of *ibp* on pBTs1, however, did not show any resemblance to a σ^{32} promoter sequence, an observation similar to those made for some other hsp20 genes, including *Bradyrhizobium hspA* and *hspB* (50) and *Clostridium hsp18* (59).

Inspection of the regions of both plasmids where termination signals might be expected revealed, as noted before, the presence of numerous short direct and inverted repeats. The majority of these probably arose by chance, and identification

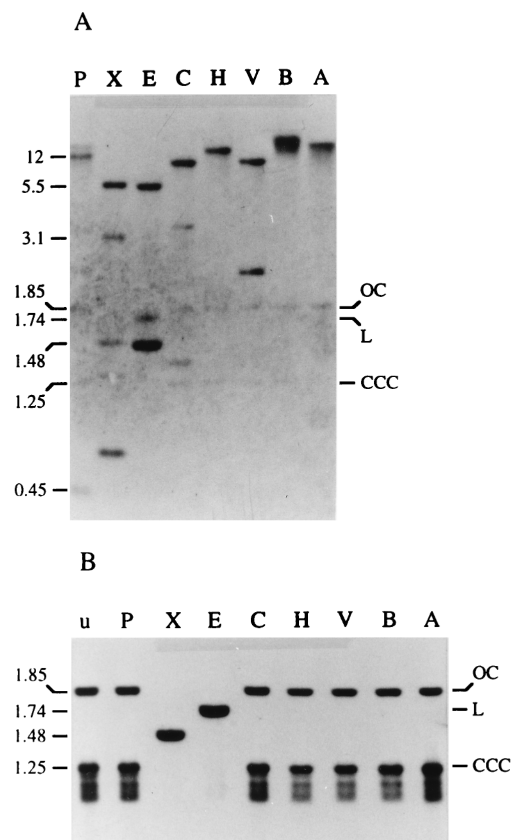

FIG. 5. Southern blot analysis of *B*(*Tetraneura caerulescens*) DNA. (A) Restriction enzyme digests of genomic DNA hybridized with probe il7.5; (B) undigested plasmid DNA and restriction enzyme digests of plasmid DNA hybridized with a probe specific to *repA*$_1$. A small fragment of 260 bp expected in the *Xba*I digest was not detected, probably due to stringency conditions used in the procedure. Lane u contained undigested plasmid DNA. P, *Pst*I; X, *Xba*I; E, *Eco*RI; C, *Cla*I; H, *Hin*dIII; V, *Eco*RV; B, *Bam*HI; A, *Acc*I. The different forms of the putative plasmid are open circular (OC), linear (L), and covalently closed circular (CCC).

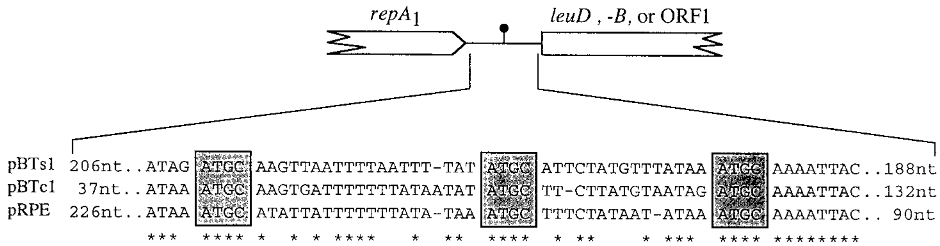

FIG. 6. Nucleotide sequence comparison of the regions downstream of $repA_1$ of pBTs1, pBTc1, and pRPE. Different genes are located at the 3′ end of the region in the three plasmids. The circle corresponds to the circle in Fig. 2.

of sites in these DNAs that could represent rho-independent termination signals is therefore very difficult.

Putative origin of replication. In order to identify regions that might function as the origins of replication of the present group of plasmids, the intergenic regions from pBTs1, pBTc1, and pRPE longer than 100 bp were pairwise compared in both orientations and inspected for stretches of similarity. Only the regions downstream of the $repA_1$ genes were found to share a significantly conserved sequence (Fig. 6). It was 56 to 58 nucleotides in length, 70% of the positions were conserved among the three plasmids, and it contained three directly repeated 5′-ATGC-3′ sequences, separated by 19 or 20 bp and 15 or 16 bp, respectively. A search for this tetranucleotide in all intergenic spacers of the three plasmids, which together sum up to 2,708 bp, revealed a total of 18 occurrences of the 5′-ATGC-3′ sequence. Nine of these were located in the identified site downstream of $repA_1$, the sequences of which sum up to only 172 bp for the three plasmids together. In addition, inspection of the same region from a related plasmid found in *B*(*Tuberolachnus salignus*), which is currently being sequenced, also revealed the presence of this sequence, including the three ATGC boxes. In the absence of any similarity beyond mere high AT content between other intergenic regions of the plasmids, these findings strongly suggest that the conservation of the region downstream of $repA_1$ reflects functional constraints and therefore seems to be the best candidate to accommodate the origin of replication.

Phylogenetic analysis of 16S rDNA sequences. In order to allow for evolutionary interpretation of the diversity in genetic organization of the leucine gene cluster, the phylogenetic relationships of the *B. aphidicola* isolates used in this study were assessed by sequence analysis of the 16S rDNA gene, taking advantage of the large existing database of sequences for this gene in other *B. aphidicola* strains (44, 49). The parameter (alpha) that describes among-site rate variation within genes was estimated by the method of Yang (67), using partially resolved trees that were based on the 16S rDNA phylogenies of Moran et al. (44) and the Ribosomal Database Project RDP (41). Alpha was approximately 0.2, which is consistent with the long-standing notion of marked among-site rate variation in rRNA genes (40, 56). A phylogenetic tree was constructed by using maximum likelihood (22), with estimated nucleotide substitution rates categorized into eight classes. The tree is presented in Fig. 7. *B*(*Thelaxes suberi*) was basal to a group comprised of *B*(*Chaitophorus viminalis*) and *B*(*Mindarus kinseyi*). *B*(*Tetraneura caerulescens*) clustered with *B. aphidicola* from aphids of the family Pemphigidae. Both phylogenetic groupings are in accordance with established classifications and the phylogeny of aphids proposed by Heie (28). At a higher level, the 16S rDNA phylogenetic relationships presented here are

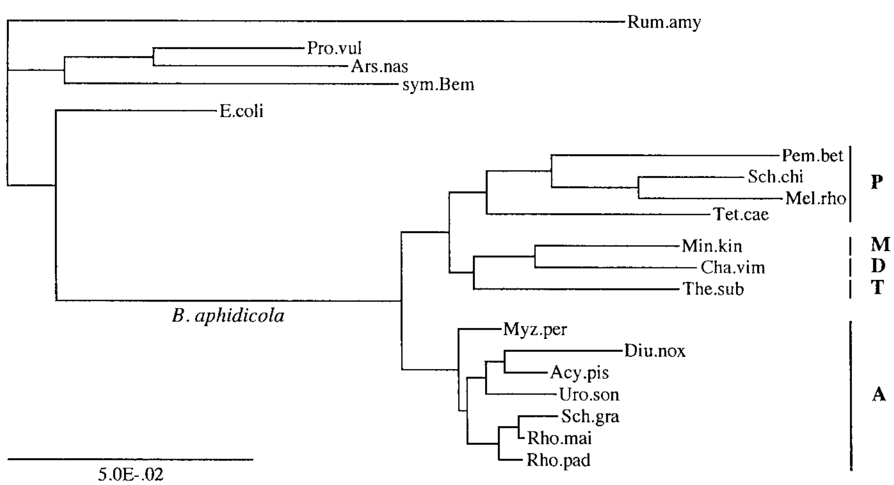

FIG. 7. Maximum likelihood tree estimated from 16S rDNA sequences of *B. aphidicola* strains from different aphid species and other bacteria selected for the outgroup. Sequences from *B*(*Thelaxes suberi*) and *B*(*Tetraneura caerulescens*) represent new data. Branch lengths are scaled according to the expected number of substitutions per site (23). The following bacteria are included: *Ruminobacter amylophilus* (Rum.amy); *Proteus vulgaris* (Pro.vul); *Arsenophonus nasoniae* (Ars.nas); symbiont of *Bemisia tabaci* (sym.Bem); *E. coli*; and *B. aphidicola* strains from the aphids *Pemphigus betae* (Pem.bet), *Schlechtendalia chinensis* (Sch.chi), *Melaphis rhois* (Mel.rho), *Tetraneura caerulescens* (Tet.cae), *Mindarus kinseyi* (Min.kic), *Chaitophorus viminalis* (Cha.vim), *Thelaxes suberi* (The.sub), *Myzus persicae* (Myz.per), *Diuraphis noxia* (Diu.nox), *Acyrthosiphon pisum* (Acy.pis), *Uroleucon sonchi* (Uro.son), *Schizaphis graminum* (Sch.gra), *Rhopalosiphum maidis* (Rho.mai), and *R. padi* (Rho.pad). Aphid families Pemphigidae (P), Mindaridae (M), Drepanosiphinidae (D), Thelaxidae (T), and Aphididae (A) are indicated.

congruent with the phylogeny of *B. aphidicola* constructed at the RDP but bear a single, though remarkable, difference from the results obtained by Munson et al. (49) and Moran et al. (44) for the same gene. Our phylogeny tree suggests an early divergence between *B. aphidicola* strains from the Aphididae and *B. aphidicola* strains from all other aphid families sampled to date, while the previous phylogenies indicated such an early divergence between *B. aphidicola* strains from the Pemphigidae and those from all other aphid families sampled.

DISCUSSION

Physiological and metabolic studies have provided evidence for essential amino acid biosynthesis by *B. aphidicola* and provisioning to the aphid host (16, 20, 58). More recently, these findings have been supported by the discovery of genetic modifications in *B. aphidicola* in the biosynthetic pathways leading to tryptophan (36) and leucine (8). However, with the notable exception of a characterization of the genes involved in tryptophan biosynthesis in *B. aphidicola* from an aphid of the family Pemphigidae (37), species from the Aphididae and their endosymbionts have almost exclusively been the subjects of study (reviewed in reference 6). For a better understanding of the evolution of the nutritional interactions and the potentials for adaptation through genome reorganization of an endosymbiotic bacterium, knowledge of the genetic diversity of *B. aphidicola* is a prerequisite. We have undertaken a comprehensive survey of the organization of the leucine gene cluster among major evolutionary lineages of *B. aphidicola*. Here, we present the first report on plasmids and genetic modifications in biosynthetic pathways of essential amino acids found in *B. aphidicola* from aphids that belong to families other than the Aphididae, i.e., Thelaxidae and Pemphigidae. Their phylogenetic affiliations were corroborated by comparative sequence analysis of the 16S rRNA genes (Fig. 7). Both families, and consequently their endosymbionts, diverged from each other and from the Aphididae at least 80 million years ago (44).

B(*Tetraneura caerulescens*) was found to contain a plasmid, designated pBTc1, that bears strong resemblance to plasmids that carry the leucine gene cluster in *B. aphidicola* strains from other aphids, i.e., pRPE, previously found in *B*(*R. padi*) (8), and pBTs1 from *B*(*Thelaxes suberi*), which is described in this study. First and foremost, the three plasmids share homologous regions that are most likely involved in replication functions. They encode either one or two variants of a replication initiation protein, which, at the amino acid sequence level, appears to be related to the RepA protein of IncFII plasmids (8) (Table 1). The presence of a highly conserved sequence of 56 to 58 nucleotides downstream of the *repA*$_1$ gene suggests that this region is under functional constraint and may contain the origin of replication of the plasmids. The sequence is further characterized by a very distinct occurrence of three ATGC boxes (Fig. 6), the function of which is yet unknown. Further study is required to confirm whether this region is essential for replication. Interestingly, however, a location of the origin of replication (*oriR*) downstream of *repA* is also a characteristic of the R-like plasmids of the IncFII incompatibility group and is physically related to the exclusive *cis* action of RepA (33). Additional features of *oriR* include (i) the absence of prominent iterated elements, which are otherwise common to origins of replication, and (ii) a DnaA box that precedes a 90-bp segment that comprises RepA binding sites and a 70-bp AT-rich region that contains an integration host factor-binding site (33). DnaA boxes that perfectly match the proposed consensus sequence have been identified on a *B. aphidicola* plasmid carrying *trpEG* (36), indicating that this element is potentially conserved in *Buchnera*. However, neither of the elements characteristic of *oriR* could be identified on the *B. aphidicola* plasmids.

The presence of two variants of the *repA* gene on pRPE and pBTs1 as opposed to the presence of only *repA*$_1$ on pBTc1 (Fig. 2A) suggests that RepA$_1$ is sufficient for replication initiation. RepA$_2$ may fulfill a dispensable function, e.g., increase of replication initiation rate resulting in higher copy number. However, the existence of two variants raises the question of whether *repA* has been duplicated on pRPE and pBTs1 or whether an ancestral replicon already contained both variants, one of which was lost from pBTc1. Unfortunately, phylogenetic analyses failed to give an answer to this question. Topologies of RepA derived trees were found to be extremely sensitive to both the method of tree construction employed and the residues included in the analyses. Further taxonomic sampling of *repA* sequences may overcome this problem.

The second feature shared by the three plasmids is the presence of an open reading frame the predicted product of which is homologous to putative integral membrane proteins of *E. coli* and *H. influenzae*, YqhA and HI0507, respectively. The proteins are highly conserved at the amino acid level, which may be indicative of strong selective constraint on their structure and/or function. The chromosomal locations of the genes in *E. coli* and *H. influenzae* suggest that they are not involved in typically plasmid-associated functions, like replication or plasmid maintenance. Yet, in widely divergent *B. aphidicola* the open reading frame is amplified on a plasmid, a kind of genetic modification that, to date, is otherwise known only for genes involved in essential amino acid biosynthesis (6, 8). In view of *B. aphidicola*'s unusual activity of essential amino acid overproduction, which is directed towards the aphid host, the assumption that active export systems are involved in the provisioning seems plausible. Several transport systems, most of which are involved in resistance against toxic compounds, have been found to be encoded by plasmids (60). It is tempting to speculate that the product of ORF1 plays a role in an extrusion system that is functionally related to the overproduction of leucine.

Ibp, encoded on pBTs1, was found to be significantly similar to members of the hsp20 family of small heat shock proteins. To our knowledge, we report here the first case of a plasmid-encoded member of this family. The small heat shock proteins constitute a ubiquitous but diverse family of proteins of 15 to 30 kDa, which share the characteristic of stress-inducible expression of genes and the tendency to form large aggregates. Chaperoning, related to particular protein metabolism, has been proposed to be the primordial and widely retained function of the family (10). The *E. coli* IbpA and -B proteins, which are most similar to the *B*(*Thelaxes suberi*) Ibp, may conform to this function (1). Several studies have indicated that intracellular bacteria of eukaryotes, pathogenic as well as symbiotic, have elevated levels of the heat shock protein GroEL (6). At least in pathogenic bacteria, this is thought to be a stress response to the hostile environment of the host. Overexpression of GroEL has also been observed in *B. aphidicola* (5), and the importance of a highly abundant chaperonin to the symbiosis has been the subject of much speculation (24, 43, 45). Our finding of a putative heat shock protein that is amplified on a plasmid may, at most, render further support to the notion that *B. aphidicola* has a relatively high need for molecular chaperones, but elucidation of specific functions awaits experimental study.

Taken together, the similarities among the three plasmids outlined above indicate that they have evolved from a common ancestor. The small plasmid pBTc1 may resemble most closely the ancestral form of a replicon that has been recruited for amplification of the *leuABCD* gene cluster, as is illustrated by pRPE and pBTs1. In *B. aphidicola* from at least one member of the Pemphigidae (*Tetraneura caerulescens*), the *leuABCD* genes are not linked to this replicon but still reside in the chromosome. We observed a similar, chromosomal organization in *B. aphidicola* from an aphid of the Drepanosiphinidae (63), which are the closest relatives of *Thelaxes suberi* (Fig. 7). Under the assumption that plasmids are not exchanged between *B. aphidicola* strains from different aphids, the phylogenetic distribution of plasmid-borne *leuABCD* genes is most parsimoniously explained by two independent transfers to a replicon resembling pBTc1 in lineages leading to *B*(*Thelaxes suberi*) and *B*(*R. padi*).

The order of the genes in the *leuABCD* cluster is notably different between pRPE and pBTs1 (Fig. 2A). The former has an order identical to that of *E. coli*, and the genes most likely form a single transcription unit in view of the close proximity of inferred start and stop codons. pBTs1 revealed the unusual gene order of *leuBCDA*, with a maximal intergenic spacing (88 bp) between *leuD* and *leuA*. On pRPE, *leuD* is immediately followed by a rho-independent termination-like signal, consisting of on an inverted repeat of 30 bp that is continuous with a run of six consecutive T's. A similar putative termination signal was not present in the region between *leuD* and *leuA* of pBTs1. However, irrespective of whether the *leu* genes of both plasmids constitute single transcription units, the difference in gene order raises the question about the ancestral organization of the cluster in endosymbiotic *B. aphidicola*. Completion of our characterization of the chromosomally borne gene cluster in *B. aphidicola* strains from diverse aphid species will allow for a proper discussion of this important issue.

In several bacteria, expression of the leucine operon is known to be controlled by intracellular concentrations of leucine through transcription attenuation (25, 26, 61, 64). Bracho et al. (8) identified a small open reading frame (31 codons) contained entirely within ORF1 of pRPE, on the same strand but in a different reading frame which comprised 12 leucine codons. As a potentially transcribed RNA from this region resembled the leader RNAs typically found at attenuator sites, it was hypothesized that the expression of both the leucine gene cluster and *repA*$_2$ is regulated by transcription attenuation. However, a similar small open reading frame rich in leucine codons was found neither within ORF1 nor elsewhere on the sequence of pBTs1, and even though a transcription attenuation mechanism may evolutionarily be under construction on pRPE, there are no indications of a potential role in regulation of expression of the leucine genes of pBTs1.

An important question concerning the phylogenetic distribution of *leuABCD*-encoding plasmids addresses the nature of the selective pressure for enhanced capacity of leucine production in *Buchnera*. Differences in growth and reproductive rates between aphid hosts have been invoked to explain the distribution of plasmid versus chromosomal locations of the *trpEG* genes (37). To date, *trpEG* amplification is known to occur only in *B. aphidicola* from aphids of the Aphididae (6). These are renowned for high growth rates (4), and the vast majority (88%) of the species thrives on herbs, including a number of agriculturally important crops (7). Species from all other aphid families may generally have slower growth rates, and typically live on trees (72%). In contrast to *trpEG*, the *leuABCD* amplification is now found not to be restricted to *B. aphidicola* from Aphididae. Unfortunately, it is not yet possible to relate this finding explicitly to growth and reproductive rates, because no data on these traits are available for *Thelaxes*. However, besides growth kinetics as a potential component of selective pressure for gene amplification and its maintenance, the potential role of the host plant deserves attention. Phloem sap composition in divergent plant species might differ enough so as to raise barriers to the aphid-*Buchnera* system that can be overcome only by adaptation of their essential amino acid biosynthesis and provisioning pathways.

ACKNOWLEDGMENTS

We are indebted to Fernando González-Candelas and the Servei de Bioinformàtica (Universitat de València) for providing computing facilities and support. We gratefully thank J. M. Michelena for identification of the aphid species.

This work was supported by an EU HCM research grant (CHRX-CT94-0660) to R.C.H.J.V. and grant PB93-0690 from DGICYT (MEC, Spain) to A.M.

REFERENCES

1. **Allen, S. P., J. O. Polazzi, J. K. Gierse, and A. M. Easton.** 1992. Two novel heat shock genes encoding proteins in response to heterologous protein expression in *Escherichia coli*. J. Bacteriol. **174:**6938–6947.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. Greene Publishing, Brooklyn, N.Y.
4. **Baumann, L., and P. Baumann.** 1994. Growth kinetics of the endosymbiont *Buchnera aphidicola* in the aphid *Schizaphis graminium*. Appl. Environ. Microbiol. **60:**3440–3443.
5. **Baumann, P., L. Baumann, and M. A. Clark.** 1996. Levels of *Buchnera aphidicola* chaperonin GroEL during growth of the aphid *Schizaphis graminum*. Curr. Microbiol. **32:**1–7.
6. **Baumann, P., L. Baumann, C. Y. Lai, D. Rouhbakhsh, N. A. Moran, and M. A. Clark.** 1995. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. Annu. Rev. Microbiol. **49:**55–94.
7. **Blackman, R. L., and V. F. Eastop.** 1984. Aphids on the world's crops. Wiley, Chichester, Great Britain.
8. **Bracho, A. M., D. Martinez-Torres, A. Moya, and A. Latorre.** 1995. Discovery and molecular characterization of a plasmid localized in *Buchnera* sp. bacterial endosymbiont of the aphid *Rhopalosiphum padi*. J. Mol. Evol. **41:**67–73.
9. **Brough, C. N., and A. F. G. Dixon.** 1990. Ultrastructural features of egg development in oviparae of the vetch aphid, *Megoura viciae* Buckton. Tissue Cell **22:**51–63.
10. **Caspers, G.-J., J. A. M. Leunissen, and W. W. de Jong.** 1995. The expanding small heat-shock protein family, and structure predictions of the conserved "α-crystallin domain." J. Mol. Evol. **40:**238–248.
11. **Chuang, S.-E., V. Burland, G. Plunkett, D. L. Daniels, and F. R. Blattner.** 1993. Sequence analysis of four heat-shock genes constituting the hslTS/ibpAB and hslVU operons in *Escherichia coli*. Gene **134:**1–6.
12. **Clark, M. A., L. Baumann, and P. Baumann.** 1992. Sequence analysis of an aphid endosymbiont DNA fragment containing *rpoD* (β-subunit of RNA polymerase) and portions of *rplL* and *rpoC*. Curr. Microbiol. **25:**283–290.
13. **Cowing, D. W., J. C. A. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix, and C. A. Gross.** 1985. Consensus sequence for *Escherichia coli* heat shock gene promoters. Proc. Natl. Acad. Sci. USA **82:**2679–2683.
14. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
15. **Douglas, A. E.** 1989. Mycetocyte symbiosis in insects. Biol. Rev. **69:**409–434.
16. **Douglas, A. E.** 1990. Nutritional interactions between *Myzus persicae* and its symbionts, p. 319–327. *In* R. K. Campbell and R. D. Eikenbary (ed.), Aphid-plant genotype interactions. Elsevier, Amsterdam, The Netherlands.
17. **Douglas, A. E.** 1992. Requirement of pea aphids (*Acyrtosiphon pisum*) for their symbiotic bacteria. Entomol. Exp. Appl. **65:**195–198.
18. **Douglas, A. E.** 1993. The nutritional quality of phloem sap utilized by natural aphid populations. Ecol. Entomol. **18:**31–38.
19. **Douglas, A. E.** 1996. Reproductive failure and the amino acid pools in pea aphids (*Acyrthosiphon pisum*) lacking symbiotic bacteria. J. Insect Physiol. **42:**247–255.
20. **Douglas, A. E., and W. A. Prosser.** 1992. Synthesis of the essential amino acid tryptophan in pea aphid (*Acyrthosiphon pisum*) symbiosis. J. Insect Physiol. **38:**565–568.
21. **Febvay, G., I. Liadouze, J. Guillaud, and G. Bonnot.** 1995. Analysis of energetic amino acid metabolism in *Acyrthosiphon pisum*—a multidimensional approach to amino acid metabolism in aphids. Arch. Insect Biochem. Physiol. **29:**45–69.
22. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum

likelihood approach. J. Mol. Evol. **17:**368–376.

23. **Felsenstein, J.** 1993. Phylogenetic inference package (PHILIP) version 3.5. University of Washington, Seattle.
24. **Fukatsu, T., and H. Ishikawa.** 1993. Occurrence of chaperonin 60 and chaperonin 10 in primary and secondary bacterial endosymbionts of aphids: implications for the evolution of an endosymbiotic system in aphids. J. Mol. Evol. **36:**568–577.
25. **Gemmill, R. M., S. R. Wessler, E. B. Keller, and J. M. Calvo.** 1979. Leu operon of *Salmonella typhimurium* is controlled by an attenuation mechanism. Proc. Natl. Acad. Sci. USA **76:**4941–4945.
26. **Godon, J.-J., M.-C. Chopin, and S. D. Ehrlich.** 1992. Branched-chain amino acid biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. J. Bacteriol. **174:** 6580–6589.
27. **Harrison, C. P., A. E. Douglas, and A. F. G. Dixon.** 1989. A rapid method to isolate symbiotic bacteria from aphids. J. Invertebr. Pathol. **53:**427–428.
28. **Heie, O. E.** 1987. Paleontology and phylogeny. World Crop Pests **2a:**367–391.
29. **Hinde, R.** 1971. Maintenance of aphid cells and the intracellular symbiotes of aphids in vitro. J. Invertebr. Pathol. **17:**333–338.
30. **Hofmann, K., and W. Stoffel.** 1993. TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler **347:**166.
31. **Houk, E. J., and G. W. Griffiths.** 1980. Intracellular symbiotes of the Homoptera. Annu. Rev. Entomol. **25:**161–187.
32. **Ishikawa, H.** 1987. Nucleotide composition and kinetic complexity of the genomic DNA of an intracellular symbiont of the pea aphid *Acyrthosiphon pisum*. J. Mol. Evol. **24:**205–211.
33. **Kornberg, A., and T. A. Baker.** 1992. DNA replication, 2nd ed. W. H. Freeman and Company, New York, N.Y.
34. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
35. **Lai, C.-Y., and P. Baumann.** 1992. Genetic analysis of an aphid endosymbiont DNA fragment homologous to the rnpA-rpmH-dnaA-dnaN-gyrB region of eubacteria. Gene **113:**175–181.
36. **Lai, C.-Y., L. Baumann, and P. Baumann.** 1994. Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. Proc. Natl. Acad. Sci. USA **91:**3819–3823.
37. **Lai, C.-Y., P. Baumann, and N. A. Moran.** 1995. Genetics of the tryptophan biosynthetic pathway of the prokaryotic endosymbiont (*Buchnera*) of the aphid *Schlechtendalia chinensis*. Insect Mol. Biol. **4:**47–59.
38. **Latorre, A., A. Moya, and J. F. Ayala.** 1986. Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc. Natl. Acad. Sci. USA **83:**8649–8653.
39. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:**1507–1516.
40. **Ludwig, W., and K. H. Schleifer.** 1994. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol. Rev. **15:**155–173.
41. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The ribosomal database project (RDP). Nucleic Acids Res. **24:**82–85.
42. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1992. Construction of T-vectors, a rapid general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154.
43. **Moran, N. A.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:**2873–2878.
44. **Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa.** 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. Ser. B **253:**167–171.
45. **Morioka, M., and H. Ishikawa.** 1992. Mutualism based on stress: selective synthesis and phosphorylation of a stress protein by an intracellular symbiont. J. Biochem. **111:**431–435.
46. **Munson, M. A., and P. Baumann.** 1993. Molecular cloning and nucleotide sequence analysis of a putative *trpDC*(*F*)*BA* operon in *Buchnera aphidicola* (endosymbiont of the aphid *Schizaphis graminum*). J. Bacteriol. **175:**6426–6432.
47. **Munson, M. A., P. Baumann, and M. G. Kinsey.** 1991. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. Int. J. Syst. Bacteriol. **41:**566–568.
48. **Munson, M. A., L. Baumann, and P. Baumann.** 1993. *Buchnera aphidicola* (a prokaryotic endosymbiont of aphids) contains a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene: sequence determination, and promoter and terminator analysis. Gene **137:**171–178.
49. **Munson, M. A., P. Baumann, M. A. Clark, L. Baumann, and N. A. Moran.** 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. **173:**6321–6324.
50. **Narberhaus, F., W. Weiglhofer, H.-M. Fischer, and H. Hennecke.** 1996. The *Bradyrhizobium japonicim* $rpoH_1$ gene encoding a σ^{32}-like protein is part of a unique heat shock gene cluster together with $groESL_1$ and three small heat shock genes. J. Bacteriol. **178:**5337–5346.
51. **Ochman, H., W. J. Ajioka, D. Garza, and D. Hartl.** 1989. Inverse polymerase chain reaction, p. 105–111. *In* H. A. Ehrlich (ed.), PCR technology: principles and applications for DNA amplification. Stockton Press, New York, N.Y.
52. **Ohtaka, C., H. Nakamura, and H. Ishikawa.** 1992. Structures of chaperonins from an intracellular symbiont and their structural expression in *Escherichia coli groE* mutants. J. Bacteriol. **174:**1869–1874.
53. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41–48.
54. **Raven, J. A.** 1983. Phytophages of xylem and phloem: a comparison of plant sap-feeders. Adv. Ecol. Res. **13:**136–234.
55. **Roubakhsh, D., and P. Baumann.** 1995. Characterization of a putative 23S-5S rRNA operon of *Buchnera aphidicola* (a prokaryotic endosymbiont of aphids) unlinked to the 16S rRNA-encoding gene. Gene **155:**107–112.
56. **Rzhetsky, A.** 1995. Estimating substitution rates in ribosomal RNA genes. Genetics **141:**771–783.
57. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
58. **Sasaki, T., H. Hayashi, and H. Ishikawa.** 1991. Growth and reproduction of symbiotic and aposymbiotic pea aphids, *Acyrtosiphon pisum*, maintained on artificial diets. J. Insect Physiol. **37:**85–92.
59. **Sauer, U., and P. Dürre.** 1993. Sequence and molecular characterization of a DNA region encoding a small heat shock protein of *Clostridium acetobutylicum*. J. Bacteriol. **175:**3394–3400.
60. **Tisa, L. S., and B. P. Rosen.** 1990. Transport systems encoded by bacterial plasmids. J. Bioenerg. Biomembr. **22:**493–507.
61. **Umbarger, H. E.** 1987. Biosynthesis of the branched-chain amino acids, p. 352–367. *In* F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
62. **Untermann, B. M., P. Baumann, and D. L. McLean.** 1989. Pea aphid symbiont relationships established by analysis of 16S rRNAs. J. Bacteriol. **171:** 2970–2974.
63. **van Ham, R. C. H. J., B. Sabater, and A. Latorre.** Unpublished data.
64. **Wessler, S. R., and J. M. Calvo.** 1981. Control of leu operon expression in *Escherichia coli* by a transcription attenuation mechanism. J. Mol. Biol. **149:**579–597.
65. **Womble, D. D., and R. H. Rownd.** 1988. Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. Microbiol. Rev. **52:**433–451.
66. **Yang, Z.** 1996. Phylogenetic analysis by maximum likelihood (PAML) version 1.2. Department of Integrative Biology, University of California, Berkeley.
67. **Yang, Z.** 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11:**367–372.